LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-TR-209542

Simulation of Biochemical Pathway Adaptability Using Evolutionary Algorithms

W. J. Bosl

February 9, 2005

Disclaimer

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

FY04 LDRD Final Report

Simulation of Biochemical Pathway Adaptability Using Evolutionary Algorithms

LDRD Project Tracking Code: 04-FS-008

William Bosl, Principal Investigator

Abstract

The systems approach to genomics seeks quantitative and predictive descriptions of cells and organisms. However, both the theoretical and experimental methods necessary for such studies still need to be developed. We are far from understanding even the simplest collective behavior of biomolecules, cells or organisms. A key aspect to all biological problems, including environmental microbiology, evolution of infectious diseases, and the adaptation of cancer cells is the evolvability of genomes. This is particularly important for Genomes to Life missions, which tend to focus on the prospect of engineering microorganisms to achieve desired goals in environmental remediation and climate change mitigation, and energy production. All of these will require quantitative tools for understanding the evolvability of organisms. Laboratory biodefense goals will need quantitative tools for predicting complicated host-pathogen interactions and finding counter-measures. In this project, we seek to develop methods to simulate how external and internal signals cause the genetic apparatus to adapt and organize to produce complex biochemical systems to achieve survival. This project is specifically directed toward building a computational methodology for simulating the adaptability of genomes.

This project investigated the feasibility of using a novel quantitative approach to studying the adaptability of genomes and biochemical pathways. This effort was intended to be the preliminary part of a larger, long-term effort between key leaders in computational and systems biology at Harvard University and LLNL, with Dr. Bosl as the lead PI. Scientific goals for the long-term project include the development and testing of new hypotheses to explain the observed adaptability of yeast biochemical pathways when the myosin-II gene is deleted and the development of a novel data-driven evolutionary computation as a way to connect exploratory computational simulation with hypothesis-driven experimentation. This LDRD will focus on developing prototype software for the evolutionary computation and demonstrating its efficacy on a well-known biochemical pathway in yeast.

Expected outcomes from this LDRD project included a demonstration of computational modeling of evolvability in a biochemical pathway, an important collaboration with the Systems Biology department at Harvard University, several proposals to secure external long-term funding from one or more sources and the nucleus of a new, focused research effort at LLNL in computational genomics, focused principally on Genomes to Life goals. All of these goals were achieved.

Cell Cycle Regulation and Cytokinesis Pathways

The ability of cells and organisms to generate heritable alternative strategies through genetic variations is not well understood and is a central issue in molecular biology. This issue is directly relevant to a variety of medical concerns such as cancer treatment. Cancer cells have incredible abilities to accumulate large genetic changes and evolve pathways to escape treatments that are intended to block their growth. The ability of cells to evolve alternative pathways for survival, their degree of evolutionary fitness, is sometimes referred to as evolvability(Gerhart and Kirschner 1997; Kirschner and Gerhart 1998): "the eukaryotic cell has many systems that are highly plastic and adaptable and … these systems can be reorganized to give new function" (Gerhart and Kirschner 1997). A predictive understanding of the evolutionary adaptability of cells is critical for understanding many diseases and is a central challenge for genomics research that will require new computational methods combined with new data and data analysis methods. An important research question concerns evolutionary adaptability itself: if we can learn to model the evolvability of cellular systems, it may be possible to predict unknown regulatory pathways by recapitulating them computationally.

A common problem encountered when modeling cell processes is the lack of precise knowledge of many of the parameters. We hope that one way of discovering network parameters and interactions will be to use genetic algorithms to find acceptable parameters and networks that can function as adaptations to existing (known) network models. Genetic algorithms can incorporate data other than precise numerical parameters, such as known physiologically viable ranges for values, as part of the constraints on the system. This general approach will be applied to a specific cell pathway that is intrinsically important and exhibits adaptive evolutionary abilities. An outstanding biological question in this research is to determine just how restrictive nature is when designing new pathways (or, when we attempt to discover unknown pathways from myriad possibilities).

We used this LDRD research to investigate the feasibility of needed fundamental computational methods. These will then be used to explore an important and difficult biological question concerning the cytokinesis and exit cycles in yeast. The cytokinesis pathway has shown remarkable adaptive evolution in response to deletions (Tolliday, Pitcher et al. 2003). When the myosin-II gene is deleted from the genome in budding yeast, suppressors have been observed to emerge at a fast rate and allow cells to divide in the absence of myosin. A similar high level of adaptability is observed when the Bee1 gene, a major activator of the Arp2/3 actin nucleation complex, is deleted. The genetic or epigenetic changes that cause these and other suppressors is not known, or whether they are the same changes in all cases or many different new pathways. The methods developed in this ldrd will be used to generate new hypotheses to explain these observed suppressors.

(Dunham, Badrane et al. 2002) found that exposure to a persistent new environment that lacked certain nutrients, caused new strains of budding yeast to evolve that were better able to survive in the new conditions. Genome-wide studies of gene expression revealed that independently evolved strains exhibited only a small number of possible genome expression changes. Durham concluded that the environment and regulatory networks limited the number of ways that the genome could be altered to improve fitness. If this

observation holds more generally, then we might expect that the number of new solutions available to organisms when confronted with new survival problems may be fairly tightly constrained. We would expect, then, if an evolutionary model of genetic or regulatory adaptation captures the essential features of real adaptive responses, the number of solutions generated should be rather small also. The relevance of this discovery to GtL goals is clear: we must understand and predict how natural or engineered genomes will respond to their environments if we are to realize DOE mission goals in climate, energy and environmental remediation.

Modeling Complex Systems

Fundamentally, the mitotic exit system, which includes the Cdc14 early anaphase release (FEAR) and mitotic exit network (MEN) pathways, has relatively simple functionality. Yet an intricate complex control system has evolved to make the basic functions robust under a variety of circumstances. From an engineering perspective, the mitotic exit system, like many other biological pathways, is a bona fide complex adaptive system. The key difference between complex systems and complicated, but non-complex, systems is that the former do not follow a pre-made blueprint but their design emerges from evolutionary processes (Holland 1995; Bar-Yam 2000). Traditional top-down engineering requires that all system behavior can be determined by specific components and that the behavior of the whole is precisely the sum of the parts. This is also an assumption that many biologists make implicitly in their effort to understand complicated biological processes.

An important result of evolutionary design, however, is that the dynamics of a complex system cannot be understood from its components and their interactions alone. The whole is *more than* the sum of the parts, which also imposes a natural scale on the system, below which system functions are lost. Some system functions cannot be found in any single component, but exist only when components are combined in a certain configuration. In general, evolutionary design proceeds by allowing natural selection to manipulate components to construct a (complex) system that achieves the desired global behavior. The resulting designs often look very different from those that an engineer following traditional design principles would concoct (Antonsson and Cagan 2001). It is important, however, to emphasize that there are significant differences between evolutionary algorithms applied to engineering design and the evolutionary processes that occur in biological systems. Nevertheless, both natural and artificial evolved systems exhibit properties unlike traditionally engineered systems and this perspective may help to understand large regulatory networks such as the mitotic exit control system. The purpose of this article is not to provide a comprehensive review of mitotic exit regulators and pathways (for that, several excellent recent reviews are available (Morgan 1999; Simanis 2003; Murray 2004; Seshan and Amon 2004). Instead, we focus on several important yet puzzling features of the mitotic exit system and attempt an explanation of the underlying design principles from the perspective of complex systems constructed through evolutionary processes.

The response of highly complex engineered control systems, such as modern aircraft, the space shuttle, and the national electrical supply grid, are regularly analyzed with computational models. The nonlinear response of such complex systems cannot be understood in any other way. Similarly, the mechanisms involved in many regulatory

pathways as well as pathways for structural assembly are highly complex, and computational models are being used to understand the dynamics of these systems. A key difference between engineered systems and cellular systems is evolvability. This suggests the importance of modeling the flexibility of certain important signaling or structural pathways during the cell cycle to predict the likelihood for emergence of alternative solutions. Modeling of cellular processes usually refers to quantitative rules to describe the time series of concentrations of mRNA, proteins, and other biochemicals involved in a particular pathway of interest. On different time scale, the pathway itself may change, a process referred to as evolvability. Understanding this process is important because it is how we believe complex pathways organize from simpler structures. It is also the process by which alternative pathways arise when the primary pathway is altered.

Genetic algorithms are a general class of computational methods that are inspired by natural selection and genetics. They are relatively simple to program, are highly robust, and have been applied to a wide range of engineering problems (Banzhaf 1998). In particular, they have been used to find optimal solutions to engineering systems and to generate new solutions for electrical circuit designs. This suggests that genetic algorithms might be extremely useful in conjunction with models of regulatory pathways in cells to imitate the evolution of alternative pathways. Some initial efforts to use genetic algorithms to reverse engineer simple metabolic pathways have been successful.

Data-driven Modeling and Experimental Verification

One of the appealing aspects of a genetic algorithm simulation of yeast cell cycle pathways is that they can also be investigated experimentally. Certain key components of a pathway can be deleted and the suppressors (cells that can now circumvent the requirement for this component) that arise naturally can be analyzed. Not only are microarray data, known biochemical processes and biological constraints used to drive the genetic algorithm, but laboratory experiments can also be carried out that correspond to the computational experiments. For this proposal, synthetic and microarray data will be used to demonstrate the effectiveness of our approach. Our original plan was to use a differential equation model for the pathway dynamics. During a month long visit to the Systems Biology department at Harvard University we altered this plan and decided to use a type of functional Petri net for the model.

A Method for Studying Evolutionary Adaptation in Cellular Systems

Several significant results were expected from this research. First, our specific goal is to establish that new hypotheses can be generated for unknown elements in a biochemical pathway. To do this, we will use a known mathematical model of the relevant pathways, analyze existing data or simulate synthetic data if needed, and use the model in a genetic algorithm to generate new hypotheses. The mathematical model can later be extended or modified to study more complicated models of the cytokinesis and exit cycle pathways in yeast. Perhaps most importantly, a computational method will be developed by combining existing algorithms that take into account the evolutionary adaptability of biological systems. This evolutionary modeling approach, when combined with recent and new data collection technologies, will introduce a new approach for understanding the system-level molecular response of cellular processes in a wide variety of contexts that will have direct relevance to DOE mission-critical biological questions. Figure 1 is a

schematic diagram showing the analogy between natural evolution and adaptation and the computational ideas that will be explored in this research project.

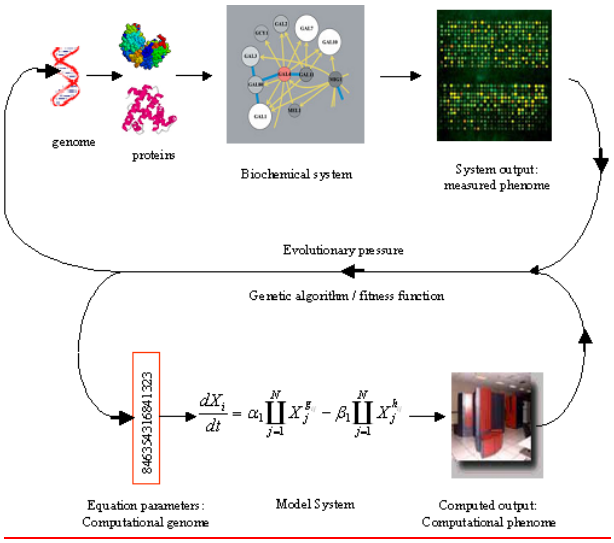

Figure 1. Evolutionary pressures on the phenotype of real organisms cause hereditable changes to the genome code, which changes the chemical properties of expressed proteins, changing the system. Analogously, the fitness values assigned to system output select for appropriate changes to the computational genome.

Research Activities and Results

The primary activities required for this project were to extract detailed information about the mitotic exit system pathway in yeast from the literature and organize it into pathway diagrams. This information served as the organizing information for the long-term research program. Figure 2 is a condensation of the results of this effort. A review paper discussing outstanding problems in the mitotic exit system has been written and submitted.

A second major effort for this project was to identify an appropriate quantitative model for the pathway. I spent a month as a visiting scientist in the Systems Biology department at Harvard Medical School in order to investigate this. Our original plan was to use differential equations, a common approach taken for modeling metabolic reactions in cells. The problem with this approach is that it requires detailed data that is not available and is not going to become available in the near future. Furthermore, the kinds of questions that biologists wish to ask about biological pathways don't require that level of detail. Rather, the questions concern connectivity of the graph, order of events (not than detailed time series), and qualitative time series for the concentrations of major reactions. A sequence of both continuous chemical reactions and discrete events characterizes the key cellular system dynamics that are of interest for this project.

With these goals in mind, we developed a modeling strategy based on a graph formalism, essentially a functional Petri net, with options for representing continuous reactions using functions (Girault and Valk 2003) or fuzzy logic (Ross 2004). A Petri net is a graphical

and mathematical modeling tool for simulating the dynamic and concurrent activities of systems. As a mathematical tool, Petri systems can be used to model equations of state, algebraic equations, and other mathematical models governing systems behavior. As a graphical tool, Petri nets can be used as a visual-communications aid similar to flow charts, block diagrams, and networks. Simulations of a simple cell cycle were carried out using the fuzzy Petri net formalism. Results are shown in figure 3. The time series produced by this approach is quite reasonable, even though only semi-quantitative data from the literature was used to drive the simulation. A paper detailing this approach will be forthcoming after more extensive and intricate modeling and validation is carried out. The goal of this project, to show the feasibility of the approach, was achieved.

The final major task for this project was to show that the pathway model could be evolved using a genetic algorithm. For this we created a target data set using the pathway model that was developed. We then created perturbed model pathways with unknown species and connections to see if we could evolve the pathway, using the target output data, and "discover" the correct unknown pathway. This was accomplished. Realistically, biological data tends to be noisy. To simulate noisy data, we added 1%, 5% and 10% white noise to the target data and carried out the evolution experiments. Again, the correct pathways were discovered even with 10% noise in the data. The results of an evolutionary search are shown in figure 4.

An important goal for this feasibility study was to use these simple preliminary results to write more extensive proposals for long term research and development in computational cell biology. Three proposals were written based on this work, two to NIH and one to the DOE/Genomes to Life program. Two proposals were funded and the third is now pending. Thus, all goals for this project were achieved.

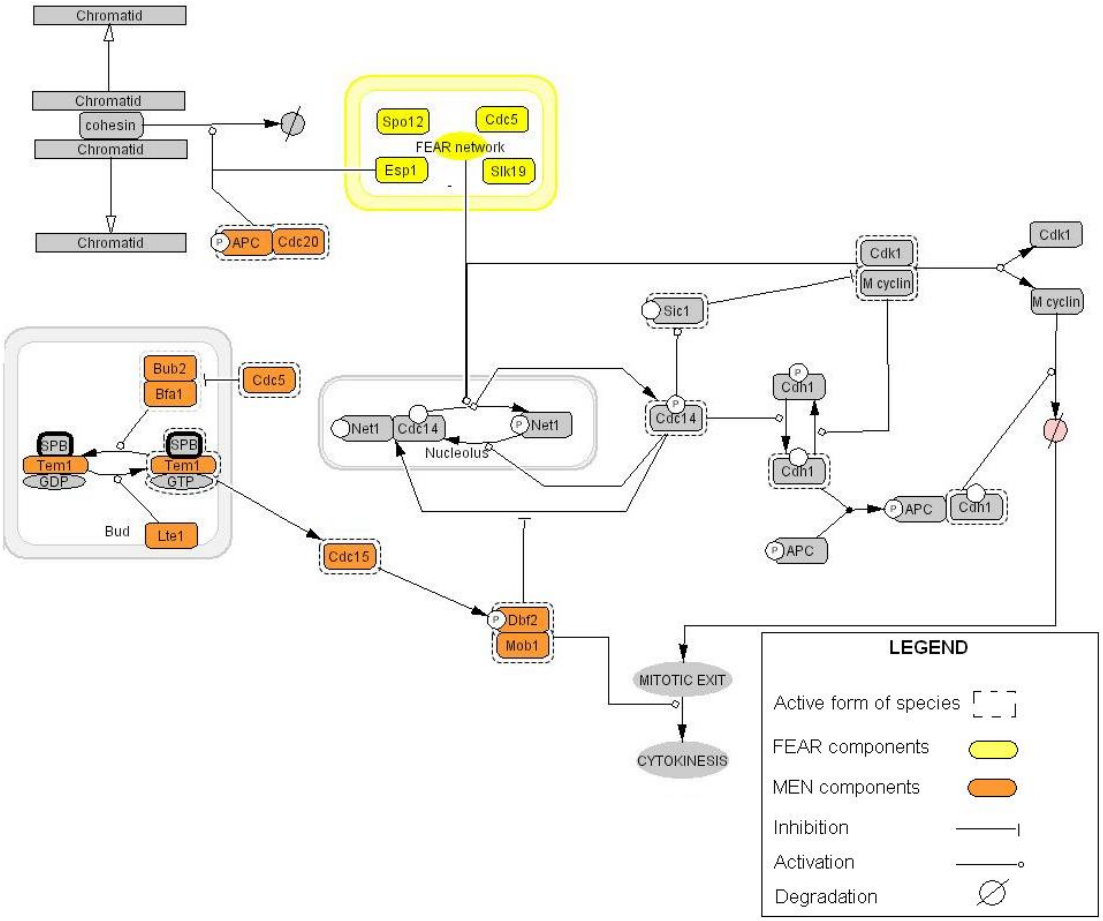

Figure 2. A simple overview of the mitotic exit system. Sister chromatid separation is initiated by activation of the APC/Cdc20 complex, which also triggers the FEAR pathway. FEAR has a dual role: it enables completion of chromatid separation and also causes transient early release of Cdc14 from its prison in the nucleolus. The MEN sustains Cdc14 release, detects proper spindle pole migration into the bud, begins the breakdown of mitotic cyclins and initiates cytokinesis, the final step in cell division.

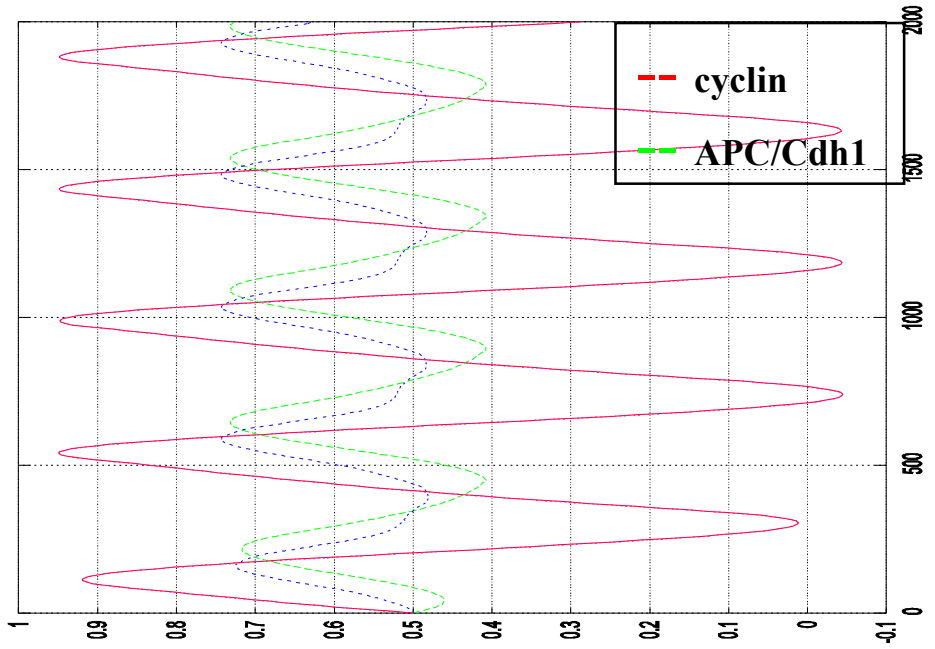

Figure 3. Four rules were used to define the dynamics of a simple cell cycle: 1. If cyclin is low/med /high, then Cdc14 is steady/inc slowly/inc rapidly; 2. If Cdc14 is low/med /high, then APC is steady/inc slowly/inc rapidly; 3. If APC is med/high, then cyclin gene is turned off (no cyclin production); 4. All species decay slowly by default. The resulting time series shows the time series that results from this rule system.

.

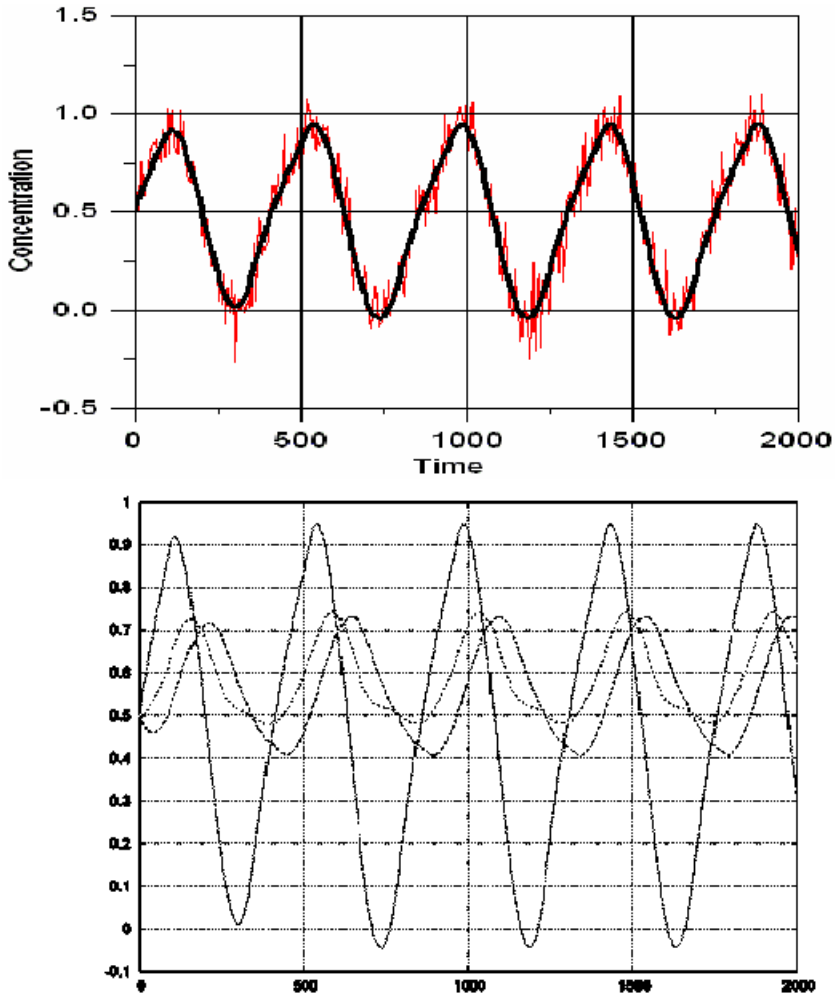

Figure 4. Noisy input data, shown in red in the top graph, was used as the fitness criterion for the genetic search algorithm. The correct reactions were discovered and gave the correct output time series for the species involved in the cell cycle, shown in the bottom graph.

Exit Plan

Several proposals have been written to enable continuation of the work started in this project. A proposal to the program in Complex Biological Systems Initiative at the National Institute of General Medical Sciences (NIGMS) of the NIH was submitted through the UC Davis Cancer Center and funded in November 2004. A proposal to the DOE Genomes to Life program through UC Merced included a research project built on this feasibility study. This project was funded and is now part of the new UC Merced Center for Computational Biology. Participation in a new LDRD project, "Characterizing the Regulatory Genome: Transcription factor proteins and gene regulation networks in living cells", and ERD LDRD project funded through BBRP was enabled in part through this feasibility study. Other proposals to NIH are currently pending.

Acknowledgements

Research collaboration from Professor Rong Li in the Cell Biology department at Harvard Medical School is gratefully acknowledged. W. Bosl thanks Rebecca Ward and Marc Kirschner for graciously inviting him to spend a month as a visiting scientist in the new Department of Systems Biology at Harvard for the month of May, 2004.

References

Antonsson, E. K. and J. Cagan (2001). Formal engineering design synthesis. Cambridge, U.K. ; New York, Cambridge University Press.

Banzhaf, W. (1998). Genetic programming : an introduction : on the automatic evolution of computer programs and its applications. San Francisco, Calif., Morgan Kaufmann.

Bar-Yam, Y. (2000). Unifying Themes in Complex Systems. Boulder, CO, Westview Press.

Dunham, M. J., H. Badrane, et al. (2002). "Characteristic genome rearrangements in experimental evolution of Saccharomyces cerevisiae." Proc Natl Acad Sci U S A 99(25): 16144-9.

Gerhart, J. and M. Kirschner (1997). Cells, embryos, and evolution : toward a cellular and developmental understanding of phenotypic variation and evolutionary adaptability. Malden, Mass., Blackwell Science.

Girault, C. and R. Valk (2003). Petri nets for systems engineering : a guide to modeling, verification, and applications. Berlin, Springer.

Holland, J. (1995). Hidden Order: How Adaptation Builds Complexity. New York, Addison-Wesley Publishing Co.

Kirschner, M. and J. Gerhart (1998). "Evolvability." Proc Natl Acad Sci U S A 95(15): 8420-7.

Morgan, D. O. (1999). "Regulation of the APC and the exit from mitosis." Nat Cell Biol 1(2): E47-53.

Murray, A. W. (2004). "Recycling the cell cycle: cyclins revisited." Cell 116(2): 221-34.

Ross, T. J. (2004). Fuzzy logic with engineering applications. Chichester, Wiley.

Seshan, A. and A. Amon (2004). "Linked for life: temporal and spatial coordination of late mitotic events." Curr Opin Cell Biol 16(1): 41-8.

Simanis, V. (2003). "The mitotic exit and septation initiation networks." J Cell Sci 116(Pt 21): 4261-2.

Tolliday, N., M. Pitcher, et al. (2003). "Direct evidence for a critical role of myosin II in budding yeast cytokinesis and the evolvability of new cytokinetic mechanisms in the absence of myosin II." Mol Biol Cell 14(2): 798-809.